1 Problemset 0 - Linear Algebra and Multivariable Calculus

1.1 Gradients and Hessians

Recall that a matrix $A \in M_{n\times n}(\mathbb{R})$ is *symmetric* if and only if $A^T = A$. Also, the gradient vector of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix}$$

Where $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$.

The hessian $\nabla^2 f(x)$ is the $n \times n$ matrix:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial^2}{\partial x_1 x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 x_n} f(x) \\ \frac{\partial^2}{\partial x_2 x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 x_n} f(x) \\ & & \cdots & \\ \frac{\partial^2}{\partial x_n x_1} f(x) & \frac{\partial^2}{\partial x_n x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n^2} f(x) \end{bmatrix}$$

Part (a): Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. Compute $\nabla f(x)$.

Solution:

Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$

By some calculations, we obtain $f(x) = \frac{1}{2}(\sum_{i=1}^n \sum_{j=1}^n a_{ij}x_ix_j) + \sum_{i=1}^n b_ix_i$

$$\implies \frac{\partial f}{\partial x_k} = \frac{1}{2}\left(\sum_{j=1}^n a_{kj}x_j + a_{jk}x_j \right) + b_k$$

(Since we remove exactly one x_k from all terms $a_{ij}x_ix_j$ such that either $i = k$ or $j = k$)
Note that $a_{kj} = a_{jk}$ since A is symmetric, we obtain:

$$\frac{\partial f}{\partial x_k} = \sum_{j=1}^n a_{kj}x_j + b_k$$

Hence:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j + b_1 \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j + b_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j \end{bmatrix} = Ax + b$$

Part (b): Suppose $f(x) = g(h(x))$, where $g : \mathbb{R} \to \mathbb{R}$ is differentiable and $h : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Compute $\nabla f(x)$ in terms of h, g.

Solution:
By the chain rule:

$$\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial h} \cdot \frac{\partial h}{\partial x_i} = g'(h) \cdot \frac{\partial h}{\partial x_i}$$

Therefore:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} = g'(h) \cdot \begin{bmatrix} \frac{\partial}{\partial x_1} h(x) \\ \frac{\partial}{\partial x_2} h(x) \\ \vdots \\ \frac{\partial}{\partial x_n} h(x) \end{bmatrix} = g'(h) \cdot \nabla h(x)$$

Part (c): Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. Compute $\nabla^2 f(x)$.

Solution:
Using **part (a)**, we have:

$$\frac{\partial f}{\partial x_i} = \sum_{j=1}^n a_{ij}x_j + b_i$$

Thus:

$$\frac{\partial^2 f}{\partial x_i x_j} = a_{ij}$$

Hence:

$$\nabla^2 f(x) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = A$$

Part (d): Let $f(x) = g(a^T x)$, where $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and $a \in \mathbb{R}^n$ is a vector. What are $\nabla f(X)$ and $\nabla^2 f(x)$?.

Solution: Let $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ and $h : \mathbb{R}^n \to \mathbb{R}$ such that $h(x) = a^T x, \forall x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$.

By the chain rule:

$$\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial h} \cdot \frac{\partial h}{\partial x_i} = g'(a^T x) a_i$$

And:

$$\frac{\partial^2 f}{\partial x_i x_j} = g''(a^T x) a_i a_j$$

Thus:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} = g'(a^T x) \cdot a$$

$$\nabla^2 f(x) = g''(a^T x) \begin{bmatrix} a_1^2 & a_1 a_2 & \cdots & a_1 a_n \\ a_2 a_1 & a_2^2 & \cdots & a_2 a_n \\ & & \cdots & \\ a_n a_1 & a_n a_2 & \cdots & a_n^2 \end{bmatrix} = g''(a^T x)(a \cdot a^T)$$

1.2 Positive definite matrix